DOCUMENT RESUME

ED 292 834 TM 011 212

AUTHOR van den Bergh, Huub
TITLE Large Scale Oracy Assessment in the Netherlands.
PUB DATE [87]
NOTE 14p.
PUB TYPE Reports - Research/Technical (143) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Classroom Research; Feasibility Studies; Grade 6; Intermediate Grades; *Measurement Techniques; *Speech Communication; Speech Skills
IDENTIFIERS Netherlands; *Oracy

ABSTRACT

As part of a feasibility assessment of the level of education in the Netherlands, oracy skills of 6th-grade students, aged 10 to 12 years, were measured. Oracy tasks were constructed from the answers of teachers, parents, and educational experts to a questionnaire on desirable skills for the age group. The six tasks developed were administered to 200 students. Taped responses were rated for content, usage, and organization. Interrater reliability was good. The majority of subjects were not unduly troubled by most of the tasks; only about 1.5% formed a problem group that failed almost all of the tasks, while 11.5% performed at a doubtful level. It was concluded that the assessment of oracy skills on a large scale is feasible in primary education. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED292834

LARGE SCALE ORACY ASSESSMENT IN THE NETHERLANDS

Huub van den Bergh

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. van den Bergh

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 011 812

BEST COPY AVAILABLE

LARGE SCALE ORACY ASSESSMENT IN THE NETHERLANDS.

Huub van den Bergh

1. Introduction.

Between 1984 and 1986 a feasibility assessment on the level of education in the Netherlands was carried out. The aim was to show that a national assessment at the end of primary education was possible. That is to say: possible in a technical sense, even for such a hard-to-measure aspect of the curriculum as the mother tongue, and acceptable for workers in the field of education.

The study was divided into four more or less independent parts: measuring pupils' skills; an inventory of the opportunities to learn in the classroom; a measure of the habits and customs relating to the language of sixth graders; and an evaluation of the acceptability of a national assessment to parents and teachers.

Part of the feasibility study was the measuring of the oracy skills of sixth grade pupils (age: 10 - 12). Thus the first thing to be done was to construct oracy tasks with which the skills of the pupils could be measured. But which objectives, what kind of tasks, could be considered desirable for sixth graders? In the feasibility study we decided on so-called functional tasks; tasks which are derived from everyday oracy situations (i.e. booking tickets for the cinema, inquiring about the train departure times etc.).

To gain an impression about the desirability of a large number of oracy situations a questionnaire was constructed. This was sent to parents, teachers and what we called 'specialists' (i.e. educational researchers, policy-makers, school psychologists, school inspectors and so on). The question they had to answer for every situation was: do you think that mastering this oracy situation is desirable for sixth graders? This resulted in a list of oracy situations which could be considered desirable.

2. Construction of tasks, the assessment and judgment of the oracy products.

From the list of desirable oracy situations, as it emerged from the educational objective questionnaire, seven situations were chosen in the first instance. These situations differed from each other in audience type (a more formal audience as opposed to a classmate), the time a child had to speak, more formal (rule-governed) as opposed to more informal situations, and the purpose of the situation (persuasion, information, inquiry, amusement, etc.).

We started with six fairly monological situations and one discussion task. Although we have developed some very attractive material for the discussion task, we failed to get the children to start a real discussion. Hence, after the pretesting of all the tasks, six remained:

1. Houses. In this task the pupil had to describe two houses from drawing of a street of eight houses in front of him. A classmate, who could see the same eight houses in a different order, had to try to recognise the two houses described.*

2. Story-retelling. The pupil had to retell a funny story he had just heard to a classmate. The original story, which was quite long and recorded on tape, concerned a strange uncle of his and was told as if the story-teller really had such a strange uncle.

3. Accident. In this task the pupil had to 'telephone the police' and report an accident that had just taken place, describing what had happened on the basis of two pictures in front of him. The role of the police was played by an assistent responding according to a prescribed standard pattern.

4. Bike. For this task the pupil had to make a telephone call, on the basis of written instructions, to a railway station bicycle hire department to inquire about cycle rental terms.

   At the end of the conversation the pupil was asked questions about the information he had just been given.

5. Spiders. Here the pupil listened to the description of a spider constructing its web. While he was listening he had to arrange in the right order six pictures showing several stages of the process. After the pupil had heard and understood the process he had to explain it to a classmate.*

6. Games. In this task the pupil learnt to play a new game. After he had learnt it and played it once with the assistant, he had to explain it to a classmate, after which they could play the game together.*

* These tasks are adaptations of tasks developed by the Assessment of Performance Unit - APU - in England

All six tasks were administered to 200 sixth grade pupils at 50 primary schools throughout the Netherlands by a trained assistant. It is beyond the scope of this paper to give a detailed description of the administration, rating procedures and results for all six tasks here. I shall therefore confine myself to one: Reporting an accident (3).

The pupils were told: 'Imagine you're standing on a very busy corner. Do you know a corner like that?' (...). `Well, you are standing there and you see a serious accident take place. Here is what you see.'

here figure 1

`Awful, isn't it? Can you tell me what you saw?'
(...). `Well, the first thing to be done is to call the police. Here's a telephone: go ahead.'

The conversation between the child and the 'police' (the assistant) was recorded on tape and rated by three experienced raters. Content, usage and organisation were all rated. Content was rated with an analytical scoring system on which every relevant detail (i.e. names of streets, number of people injured, fire on the lorry involved in the accident, etc.) was recorded, and a note was made of whether the child mentioned a content element of his own accord or after prompting by the police (e.g. 'where did you say the accident took place?'). For the usage and organisation ratings the raters could refer not only to a detailed description of both aspects but also to an example tape, containing five previously rated efforts ranging from very bad through medium to very good.

Rating with these example tapes proved to be very quick and reliable. Interrater agreement ranged from .71 to .95. At first sight, then, everything seems to be all right, but there are still a few problems to be solved. For instance, correlation between the aspects across tasks is remarkably low (ranging from .01 to .53), while the cohesion within tasks is extremely high (ranging from .47 to 1.00). Correlation within tasks was for some tasks so high, that we doubted whether different aspects of oracy were being rated (see Van Gelderen, 1987, for further details on these problems).

3. Results.

Again, I shall confine myself to the detailed description of one task, though at the end of this section a brief overview will be given of the results on all tasks.

As observed earlier, all the accident-reporting telephone conversations were rated on content (direct and indirect), usage and organisation. Tables 1 and 2 give an overview of the results.

Table 1. Percentage of correctly named content elements while reporting the accident.

Content element	Named directly	Named after question from assistant
1. Accident where happened:	99 %	0 %
2. Takstraat	37 %	24 %
3. Singel	53 %	24 %
vehicles involved:		
4. Lorry	69 %	24 %
5. Car	59 %	27 %
6. Cyclist	30 %	29 %
7. Two injured	46 %	26 %
8. Car in canal	62 %	-
9. Lorry on fire	41 %	-

Table 2. Frequency distribution for the scores on usage and organisation. Both on a seven-point scale (1 = very bad, 7 = very good).

Scale point	Usage	Organisation
1 (very bad)	2 %	2 %
2	3 %	5 %
3	13 %	19 %
4 (mediocre)	27 %	27 %
5	30 %	25 %
6	18 %	14 %
7 (very good)	7 %	8 %

As can be seen from the tables 1 and 2, reporting an accident looks a rather easy task. However, in terms of successful communication between pupil and police, at least three content elements have to be put forward: that an accident has taken place (1), where it took place (2), and that there are casualties (3). Only 16% of the pupils named these three elements directly. Even after a question from the assistant only a minority of the pupils (41%) named these elements correctly.

The usage and organisation on this task caused little trouble for most of the pupils. However, it must be borne in mind that this task is strongly structured by the questions

of the police, who are at least partly responsible for the logical course of the conversation.

For all six tasks we laid down certain standards, as we did for the 'accident'. Our starting point was always whether the communication would be successful. By making this analysis we are now able to give an overview of the results on all six oracy tasks at once (see table 3).

Table 3. An overview of the results on six oracy tasks in the assessment of the level of education in the Netherlands.

Criterion:	Percentage
1. No task above the minimal level	.5 %
2. One or two tasks above the minimal level	1.0 %
3. Three or four tasks above the minimal level	11.5 %
4. One or no tasks below the minimal level	87.0 %

What can be learned from table 3? First: there are only very few minimal speakers. Group 1 and group 2 together constitute a 'problem-group' of about 1.5% ; pupils who 'fail' at almost every task. Second: the group of speakers who performed at a doubtful level is about 11.5% . Last: the majority of the pupils were not unduly troubled by most of the oral tasks.

4. The curriculum.

Pupil performance is only part of the level of education. Another part is the curriculum: what is the attitude to oracy skills in the classroom? To gain an impression of the oracy curriculum a questionnaire was sent to about 500 primary schools. The results can be summarized as follows: the mean time spent on oracy exercises, discussions etc. is 1½ hours a week, consisting of about seven classroom discussions and/or conversations of at least ten minutes. Most of the time the teacher acts as the leader of the conversation. In most of these conversations and discussions neither oracy skills nor listening skills are the main objective. Instead, they are

used only as a way of working, of pursuing some other objective such as learning from each other, learning how to get on with each other, the development of spontaneity and the creation of a pleasant atmosphere in the class.

Oracy tasks, as used in the feasibility study, are not all widely taught. Asking for or receiving information is, not surprisingly, hardly practised at all. Explaining or describing something is only sometimes practised. Only the task of retelling a story or retelling an event is quite popular. The same holds for traditional turns at speaking in front of the class, although this kind of exercise is now quite rare in modern schools.

From a questionnaire on the habits and customs related to language answered by ± 2100 sixth graders, it emerged that language subjects at schools are not very popular: all other subjects score better on appreciation. There is one exception, however, namely reading, which, after biology, is one of the most popular subjects.

We also asked the pupils how much of their spare time was spent on reading and writing, in relation to watching television. The answers made it clear that reading was a more popular pastime than writing. Only a small minority of the children said that they wrote at home for their own amusement. But even here, watching television still came top of the list.

Here are some figures: on average, 10-12-year-olds spend about 10 minutes writing a day, half an hour reading, and at least two hours watching television.

5. Conclusions.

The main conclusions from the feasibility study are: an assessment of oracy skills in primary education is technically feasible and is accepted by the teachers, parents, policymakers etc. after the feasibility study only about 2% of head teachers and 3% of parents are opposed to a periodical national assessment.

Although an assessment seems possible there are still a few hard nuts to crack. First of all there is the problem of the very low correlation of aspects across different tasks, and the high correlation of aspects within tasks. This calls for a detailed classification scheme for oracy situations, so that an adequate sample of 'situations' or tasks to be tested can be drawn up, along with a careful selection of the aspects to be judged.

One of the most important aspects of a periodical assessment of the performance of pupils in education is a comparison of performance in two (or more) assessments. This still causes considerable difficulty because none of the language and oracy tasks have proved to be scalable. Hence, a comparison can only be made by submitting the same task in two assessments, although there are a lot of undesirable side-effects to this practice.

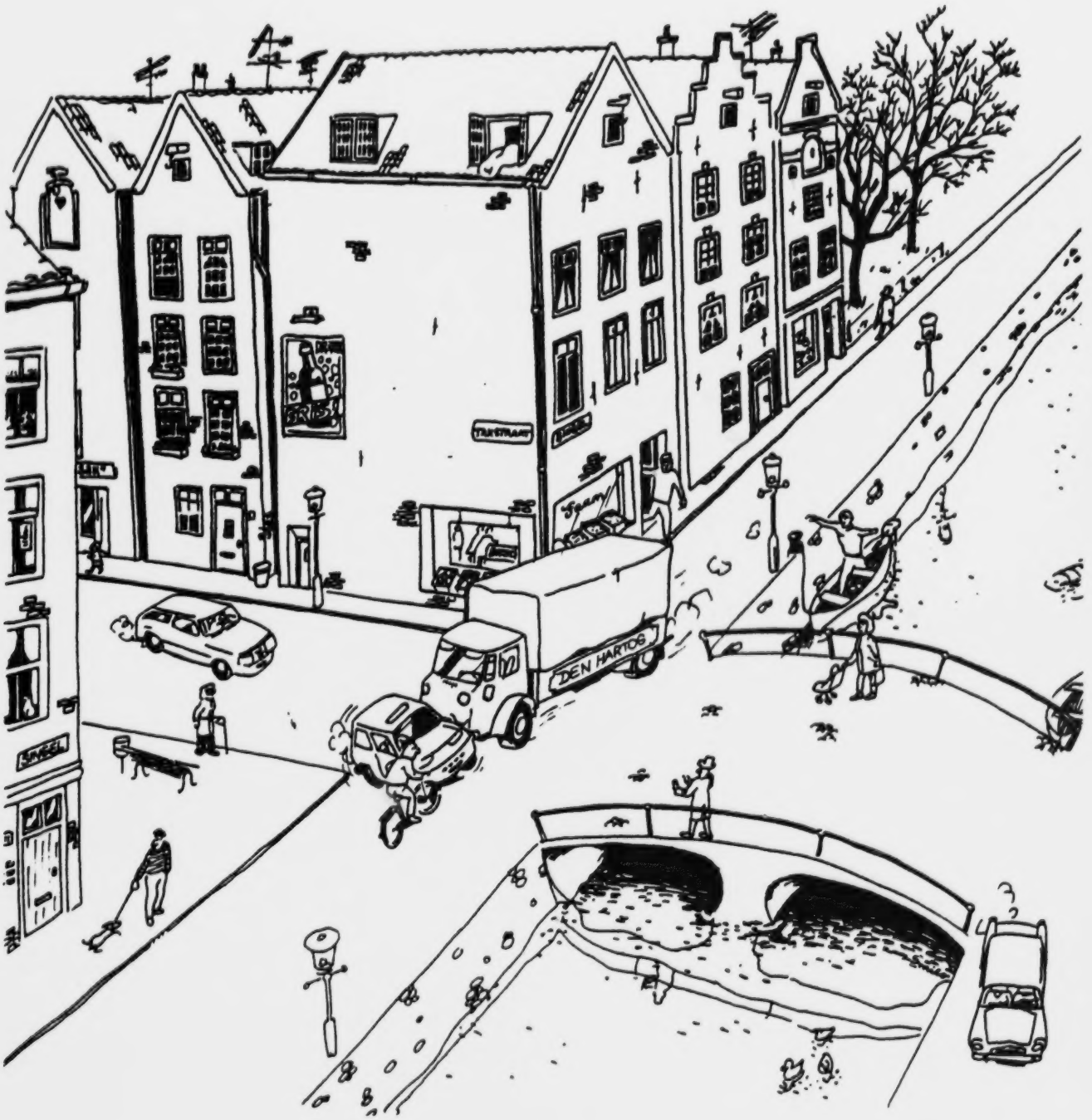
There is also a problem related to the inventorying of children's customs and habits (or attitudes). We simply asked them 'how much time did you spend last week on .... (reading, writing, television)?' But considering the answers of some pupils (e.g. pupils who watch television 12 hours a day, 8 days a week), we have doubts about the validity of this kind of research.

Literature

Bergh, H. van den, Voorstudie Periodieke Peiling van het Onderwijsniveau; deel II: Spreek- en luisterprestaties van zesdeklassers. Amsterdam, SCO, 1985.

Bergh, H. van den & J.B. Hoeksma, Functionele taaldoelen voor het lager onderwijs. SCO-cahier nr.6, Amsterdam, SCO, 1985.

Figure 1. The drawings as presented to the pupils


DEN HARTOG